



Le geste canal de communication homme / machine. Models for gestual interactions.

Claude Cadoz

► To cite this version:

Claude Cadoz. Le geste canal de communication homme / machine. Models for gestual interactions..
6th Autumn School, Sep 1992, Jouy-en-Josas, France. pp.39. hal-00910477v2

HAL Id: hal-00910477

<https://hal.science/hal-00910477v2>

Submitted on 25 Sep 2015

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

autumn SCHOOL

' 92

SEPTEMBER 7-11TH

CAMPUS THOMSON

Jouy-en-Josas

Man Machine Interaction

© Le geste canal de communication homme/machine
Models for gestual interactions

Claude Cadoz

*

ACROE

Association pour la Création et la Recherche sur les Outils d'Expression
Ministère de la Culture

LIFIA

Laboratoire d'Informatique Fondamentale et d'Intelligence Artificielle

IMAG

Institut d'Informatique et de Mathématiques Appliquées de Grenoble
46 Av. Felix Viallet, 38000 GRENOBLE - France

tel : (33) 76 57 46 61
fax : (33) 76 57 46 02
email : cadoz@lifia.imag.fr

(5 Mars 1992)

(Tous droits réservés -)
cadoz - ACROE - LIFIA -

Introduction

La Communication Homme/Machine

Le fantasme anthropomorphique à propos des machines ou peut-être le désir, vieux comme l'homme, de créer des êtres artificiels qui puissent lui venir en aide, le seconder, voir le dépasser, est complètement présent aux origines de l'ordinateur. Cependant, curieusement, si les pionniers comme Von Neumann lui-même, dès la mise en œuvre des premières machines automatiques à programmation vers 1943-45, n'ont pas hésité à parler de "cerveaux artificiels", ce n'est que beaucoup plus tard, vers la fin des années 60 que le terme de "communication homme/machine" faisant implicitement de la machine l'égal de l'homme a commencé à circuler sans d'ailleurs qu'une signification bien rigoureuse lui soit attribuée.

Au début, la machine n'est pas spécialement "communicative" ; le premier *calculateur universel contrôlé par un programme*, le Z3 de l'allemand Zuse, achevé en décembre 1941 reçoit ses données avec un lecteur de bandes et il y a une console pour l'opérateur. Pour le calculateur à relais binaires "BTL Model 1", construit en avril 1939 aux Bell Labs par les américains Stibitz et Williams, un télétype transformé servait à introduire les problèmes et à lire les résultats. Pour programmer l'ENIAC (17468 tubes à vide, 6000 commutateurs manuels), la machine qui inaugure l'ère des ordinateurs, il fallait tourner à la main des milliers de commutateurs et brancher spécialement des centaines de cables.

Le besoin de faciliter ce que l'on n'appelait donc pas encore la "communication", mais la programmation, c'est-à-dire le conditionnement des processus internes, puis la fourniture des "données d'entrée" et la "lecture des données de sortie", s'est manifesté assez vite. Dès 1948 Turing avait conçu, pour aider les "coders" (les programmeurs) du MARK 1, l'ordinateur construit par Max Newman à Manchester en Angleterre, un manuel comportant une sorte de code sténographique pour écrire des programmes, utilisant les touches d'un télétype pour représenter de longues séries de chiffres binaires. Cette machine utilisait d'ailleurs pour sa mémoire, des tubes à rayons cathodiques (CRT). Les bits étaient conservés sous la forme de points visibles sur l'écran et un CRT spécial permettait de visualiser le contenu des mémoires, ce qui constituait une première visualisation à l'écran.

Une première étape significative en matière de communication est celle du développement des "langages" de programmation et du concept corrélat de "compilateur". La forme des données et représentations binaires élémentaires prend le nom de "langage machine" et met en évidence une distance, à franchir, entre le langage de la machine et celui de l'opérateur. Le FORTRAN, mis au point sur l'IBM 701 par Backus, de 1953 à 56, est le premier "langage de communication avec la machine" qui n'est pas un langage machine et qui se veut plus proche du langage naturel.

En fait, pendant ces années mais aussi toutes celles de la "seconde informatique", des gros ordinateurs scientifiques, de l'avènement du circuit intégré, du premier développement du marché civil de l'ordinateur, de la formation de la discipline Informatique, de l'introduction de l'IA, etc., c'est-à-dire une période qui va jusque vers la fin des années 70, la notion de communication avec l'ordinateur se définit exclusivement autour de celle de langage de programmation évolué. On accorde une importance essentielle à l'aspect linguistique en quelque sorte, donc relativement abstrait et immatériel de la communication, sans se préoccuper de son mode, ou de sa *modalité*. Les organes matériels de cette communication, s'ils évoluent dans leurs performances, sont stables dans leur fonction ; il s'agit pendant très longtemps des dispositifs qui avaient déjà cours dès les années 50 : claviers alpha-numériques, consoles de

visualisation, "crayons optiques", imprimantes. Les perforateurs, lecteurs et éditeurs de cartes ou de rubans ont subsisté longtemps (pratiquement jusque vers le milieu des années 70). Les dispositifs les plus exotiques étaient les tablettes graphiques et les tables traçantes.

Au début des années 70 cependant, on s'intéressait déjà fortement dans nombre de laboratoires, aux Bell Labs aux Etats Unis par exemple, mais aussi dans des laboratoires français et européens, à Grenoble, à Paris, Stockholm, etc., à la parole comme mode de communication. Pari très ambitieux à l'époque : parler à l'ordinateur et entendre sa réponse par le même mode d'"expression". C'est une discipline de recherche en soi qui s'est alors développée : la *Communication Parlée*, avec les problèmes de l'analyse, de la reconnaissance, de l'interprétation, dans le sens de l'homme vers la machine et ceux de la construction des phrases,... de la synthèse dans l'autre. On s'aperçut au bout d'une dizaine d'années, de la très grande difficulté du domaine, de la nécessité de faire intervenir des disciplines extérieures aux sciences physiques de l'ingénieur, comme la phonétique, la psychoacoustique, la linguistique etc.

Dans ce même temps, la synthèse d'image par ordinateur, l'infographie en général, ont effectué un parcours important dont l'un des aboutissements est par exemple la performance de *réalisme* des "nouvelles images", entièrement synthétiques.

De nombreux travaux dans des domaines et pour des applications très divers : télémanipulation et commande de robots, imagerie médicale, ergonomie des postes de pilotage, des stations de contrôle de systèmes et chaînes de processus complexes, bureautique, etc., se sont développés de la même manière, sans liens systématiques les uns avec les autres, et sans se réclamer explicitement de la communication homme/machine. Le phénomène qui a provoqué alors le déclenchement d'une véritable sensibilisation à la "Communication Homme/Machine", et qui, par là même a permis de commencer à donner un sens précis au terme est probablement la mini révolution provoquée par l'arrivée du MacIntosh. Brillante vedette des PC, il apporte sans vergogne, avec la souris, les icones et les fenêtres, des concepts nouveaux et instantanément convaincants (issus de travaux et de réflexions menés antérieurement dans les labos de Xerox) et tout un mode de manipulation de représentations métaphoriques évoquant des manipulations d'objets naturels.

La communication homme/machine est devenue un thème de recherche en soi et sous la poussée d'un certain nombre de travaux, de réflexions et de nécessités, elle est actuellement en pleine structuration. On sonne le rassemblement des domaines qu'elle doit faire intervenir : la parole et les techniques de son analyse et de sa synthèse, la phonétique, la linguistique, la psychologie perceptive et cognitive, l'image et sa synthèse, la reconnaissance de forme et la vision artificielle, les techniques et théories de l'interfaçage graphique etc. Le summum est dans cette visée intégrante des "Réalités Virtuelles" très en vogue aujourd'hui, pour le plus médiatique, et dans la Communication Homme/Machine Multi-Modale pour les plus consciencieux qui envisagent la la question non seulement comme une mise à niveau des modes communicationnels imposés par la machine aux modes proprement humains mais aussi comme une articulation, à l'instar de la communication naturelle, de toutes les modalités humaines disponibles.

Mais par dessus tout cela ce qui est important, c'est le sens de l'évolution, de l'évolution technologique et de l'évolution culturelle : au départ les performances des premières machines à calculer (capables par exemple de faire la multiplication de deux nombres de plus de 20 chiffres en 3 secondes...), bien qu'inégalables par l'homme étaient en fait d'un niveau de complexité très bas comparé aux facultés humaines. Et les moyens de communication, interrupteurs et données binaires, étaient du même degré d'éléментарité que les bits et les circuits logiques de ces premiers stades du

développement de la "technologie de l'information". La progression des machines a porté la complexité des processus internes, la quantité des informations et la nécessité de leur structuration à un très haut degré. C'est cette évolution qui a amené à situer l'interaction informationnelle avec la machine sur un plan comparable à celui des phénomènes communicationnels humains. L'ordinateur tend alors à devenir une machine de représentation de phénomènes, de processus, d'entités dont la complexité interne leur permet d'atteindre la simplicité et l'évidence des objets de l'environnement humain. Dans le même temps, cette confrontation rendue possible entre les processus informatiques et les processus humains, communicationnels ou internes, fait apparaître à quel point la machine ne peut pas être comparée à l'homme. On est alors passé d'un fantasme anthropomorphique centré sur l'intérieur, le "cerveau artificiel", à une *métaphore anthropomorphique*, beaucoup plus raisonnable et utile, concernant la périphérie, la communication. Bientôt, on comprendra peut-être que la perspective la plus fascinante n'est pas, comme le préoyaient les esprits les plus enfiévrés de l'après-guerre, que la "Machine" se substitue à l'homme, le dépassant après que celui-ci lui ait transféré son intelligence, mais qu'en tant qu'outil plus universel que tous les précédents inventés par l'homme, elle intègre toutes les formes de représentations possibles en relation avec tous les modes d'action et de transformation matériels. En toute rigueur, il n'y a pas à poser le problème de la "communication" avec la machine, mais plutôt celui de l'*interaction* entre l'homme et la machine d'une part, la machine et le monde physique d'autre part, et à considérer que l'une des fonctions de la machine est la communication... entre les hommes.

C'est à cet état des choses que je voudrais maintenant aborder la question du geste.

Le geste

Dans la perspective d'"humaniser" la relation entre l'homme et l'ordinateur, on a donc battu le rappel d'un certain nombre de disciplines y compris du domaine des sciences humaines et, au plan de base, on s'est attaché aux caractéristiques objectives de chacun des *canaux de communication*, quantifiant leurs performances et modélisant leurs fonctions. On dispose aujourd'hui d'un certain nombre de connaissances sur le fonctionnement de l'oreille et les stratégies de la perception acoustique, l'organe vocal et son contrôle moteur, la vision, les relations intermodales, etc. Symétriquement, côté machine, on a perfectionné les "interfaces" : interfaces vocales capables de saisir un certain ensemble de mots prononcés oralement, de synthétiser des paroles intelligibles, interfaces de "vision artificielle" capables d'analyser certains comportements ou certaines scènes, interfaces de visualisation capables de présenter des scènes complètes dans des espaces virtuels tridimensionnels.

Pour ce qui est du geste, il est tellement omniprésent et indispensable qu'on semble avoir oublié de s'en apercevoir. Autant les canaux communicationnels de la voix, de l'ouïe et de la vision ont été parcourus, analysés, modélisés, légitimés comme voie communicationnelles, autant le canal gestuel a été sous-estimé, sous-développé. Au début il sert à manipuler les interrupteurs, contacteurs, rotateurs, curseurs, puis il se normalise et s'installe dans un carcan trop bien optimisé, dans le confort et l'universalité sage du clavier alpha-numérique ; et entre ce dernier et le développement anarchique d'une pléthore de sticks, boules, manettes et autres "syntaxeurs", il n'y a aucune théorie de la communication gestuelle. Alors que le vocal et le visuel mobilisent des contingents de chercheurs depuis une bonne vingtaine d'années, il a fallu attendre la malicieuse invasion des *souris* et la tapageuse intrusion du *gant de données*, il y a quelques années à peine, pour pouvoir commencer à parler officiellement du *geste*.

Et pourtant, le canal gestuel est peut-être le plus singulier et le plus riche des canaux de communication...

1. Le Canal Gestuel

Une particularité fondamentale du canal gestuel, qui le distingue d'emblée des autres canaux, est qu'il est deux fois double : tout d'abord il est moyen d'action sur le monde physique et moyen de communication informationnelle, ensuite, dans la seconde fonction il est à double sens, c'est-à-dire moyen d'émission et moyen de réception d'information. Ce n'est pas banal, au point que l'on pourrait penser que c'est plutôt par peur de sa trop grande richesse qu'on l'a relégué à un rôle caricatural et subalterne. Yvette Hatwell, du Laboratoire de Psychologie Expérimentale de Grenoble, a une autre explication :

" Les contraintes qui pèsent sur la vision et le toucher sont très différentes. Efficace et rapide (...), inoffensive et discrète (elle ne modifie pas les objets sur lesquels elle se porte), la vision est considérée comme une modalité noble dont l'usage est très encouragé à des fins de connaissance ou à des fins esthétiques (les musées n'exposent que des choses "à voir").

Au contraire, des interdits sévères pèsent sur le toucher. Il viennent d'abord de l'implication directe des sensibilités cutanées et mécaniques dans les fonctions sexuelles et la vie érotique, elles-mêmes réprimées. D'autre part, contrairement à la vision, le toucher n'est pas simplement "contemplatif" ; il peut transformer l'état de la source qui le stimule..." (Hatwell, 1986).

Dans cette remarque, Yvette Hatwell ne fait référence en fait qu'à l'une des fonctions du canal gestuel : celle qui correspond au sens tactile, mais même pour ce seul sens les choses apparaissent d'emblée avec leur richesse : c'est-à-dire en particulier avec l'impossibilité de dissocier action et perception. Les systèmes perceptifs sont tous des systèmes perceptivo-moteurs : il n'y a pas de vision sans acte de regarder, sans mouvements complexes volontaires ou non de l'oeil ; c'est parce que nous avons deux yeux et la mobilité corporelle que nous sommes capables de concevoir et percevoir les trois dimensions de notre espace. On sait également que certains nerfs efférents aboutissant à la cochlée modifient les caractéristiques réceptrices des cellules ciliées selon l'attitude d'attention auditive. Mais la motricité mise en oeuvre dans le toucher est très différente de celle mise en oeuvre dans la vision car la main est l'organe privilégié d'exécution des actions pratiques instrumentales sur le monde. *"La main est une structure complexe dont la fonction perceptive est imbriquée à une fonction motrice (...) : celle d'instrument de travail grâce auquel sont réalisés la saisie, le transport et la transformation des objets à des fins utilitaires."* (Hatwell, op cit).

La main a une fonction perceptive dans les actions de préhension ou d'usinage des objets. Ces actions ne pourraient réussir sans une appréciation perceptive correcte des propriétés de ces objets avant et pendant leur transformation. Yvette Hatwell fait remarquer encore que la main est capable aussi de fonctionner, comme n'importe quel système perceptif, hors d'une stricte finalité instrumentale, c'est-à-dire dans un but purement "épistémique".

Associées à la main, on peut considérer trois fonctions différentes mais complémentaires et imbriquées : une fonction *instrumentale*, c'est-à-dire d'action matérielle, modification, transformation de l'environnement, une fonction *épistémique*, c'est-à-dire de connaissance sur l'environnement et enfin une fonction que j'appellerai ici *sémiotique*, c'est-à-dire d'émission d'information à destination de l'environnement (humain ou vivant). Chacune de ces trois fonctions prise séparément fait intervenir à des degrés variables chacune des deux autres.

Considérons alors d'un peu plus près ces différentes fonctions.

1.1. La main organe de perception : le geste épistémique

La première fonction sensorielle de la main à laquelle on pense est donc le *toucher*. En fait, le sens du toucher implique beaucoup plus que la main, il implique toute l'enveloppe corporelle ainsi que les couches musculaires profondes. La surface entière du corps est pourvue de terminaisons sensibles au contact mécanique : les fibres du follicule du poil sensibles aux vibrations et les structures spécialisées. Parmi les dernières, il y a les *corpuscules de Paccini* sensibles aux déformations et aux mouvements mais dont le champ récepteur étendu et imprécis ne permet pas d'apprécier la localisation spatiale de la stimulation, et les *corpuscules de Meissner*, au contraire directement concernés par le toucher actif grâce à leur champ récepteur précis et petit.

La densité des récepteurs n'est cependant pas la même partout, elle est très importante sur la face interne des doigts, en particulier sur la dernière phalange, sorte de fovea tactile, où le pouvoir séparateur est de 2,3 mm, contre 42 mm par exemple sur le dos (Hatwell, op cit). C'est ce qui explique pour partie la fonction prédominante de la main dans le sens tactile.

Mais on s'arrête en général à cette forme élémentaire du toucher, qui nous renseigne tout au plus sur la température ou l'état de surface des objets. Or cette forme, la *perception tactile cutanée simple*, est minimale. Les mécanorécepteurs répondent en fait à des changements de pression ou de traction plutôt qu'à la pression ou à la traction elles-mêmes. Le déplacement spatial fait alors partie intégrante de la stimulation tactile et pour le toucher, la perception est indissociable d'une certaine activité motrice.

Le véritable sens du toucher est en réalité le *senstactilo-kinesthésique* ou *haptique*. Par la combinaison de stratégies de palpations, de mouvements exploratoires et de l'évolution des informations envoyées par les différentes cellules sensibles de la peau, des muscles et des jointures, on acquiert, par la perception tactilo-kinesthésique, des informations de forme¹, orientation, distance et grandeur sur les objets. La main est alors le meilleur organe perceptif tactile non seulement parce qu'elle présente la plus grande densité de récepteurs, mais aussi parce que c'est le plus mobile et qu'elle possède un très grand nombre de degrés de liberté dans cette mobilité.

Sous certaines conditions de proximité, la perception tactilo-kinesthésique permet d'accéder aux mêmes aspects spatiaux que l'oeil. Mais ce n'est pas tout encore. Un troisième degré de perception se présente, également étroitement lié aux précédents : la perception *proprioceptive*. Des récepteurs dans les couches profondes et dans les articulations², conjointement aux récepteurs vestibulaires (ceux qui informent de la direction de la pesanteur et des accélérations, situés dans l'oreille interne), nous informent en permanence sur la position des différentes parties du corps les unes par rapport aux autres et par rapport à l'espace extérieur. Associées aux perceptions précédentes, de telle sorte que l'on pourra parler de perception *tactilo-proprio-kinesthésique*, par le canal gestuel on peut être informé, sous certaines conditions de proximité et d'action motrice, de la température, de l'état de surface, de la dureté ou de la mollesse, de la forme, orientation, distance, grandeur des objets, mais également de leur

¹ La précision de la perception tactilo-kinesthésique des formes est voisine de celle de la perception visuelle.

² Sensibilité proprioceptive : celle des muscles et des articulations. La sensibilité articulaire apprécie l'ajustage correct des segments osseux. Lorsqu'elle se transmet intégralement, la pression rapproche les surfaces articulaires l'une de l'autre au maximum et supprime toute laxité ; c'est la sensibilité articulaire qui permet de rendre compte des sensations de fermeté perçues au niveau du doigt (Dorgueville 1967).

poids, de leurs structures articulatoires ou de leur propriétés de plasticité ou de déformabilité et pour finir de leurs mouvements.

De la perception "gestuelle" à la perception multi-sensorielle

Ainsi, cette forme de perception au travers de la main est-elle manifestement plus riche que ce que l'on veut bien en penser habituellement. Il y a une certaine justice à rendre au canal gestuel. Mais en fait, dans la relation avec cette machine qui prétendait par ailleurs se poser en égal de l'homme, ce n'est pas tant le canal gestuel qui est lésé que la perception tout entière. Nous avons, dans une démarche toute cartésienne, appris à considérer que l'homme avait cinq sens, et à caractériser chacun d'eux séparément pour ne considérer en général le tout que comme une somme des parties. Nous avons cru pouvoir, en conséquence, aborder les problèmes technologiques de la communication avec la machine en étudiant et réalisant séparément des dispositifs pour la voix, des dispositifs pour la vision, et, depuis peu, des dispositifs pour le geste et pour le toucher. Nous posons alors maintenant, candides, le problème de leur combinaison et nous appelons cela la communication multi-modale.

Certes, nous avons une rétine, constituée de cellules qui sont sensibles aux ondes lumineuses, certes nous avons un tympan qui réagit aux ondes acoustiques, des mécanorécepteurs qui ont une bande passante bien définie, des papilles gustatives et des cellules olfactives. Tous ces récepteurs spécialisés sont répartis en diverses régions bien définies de notre corps. Ceci doit bien nous autoriser, sur le plan technologique, à construire des organes tout aussi spécialisés pour capter et convertir spécifiquement et respectivement les phénomènes visuels, acoustiques, mécaniques etc. Oui, mais si l'on va de la périphérie (de l'homme) vers le centre, très vite on s'aperçoit que, non seulement la perception et l'action sont étroitement imbriquées, si bien que le processus de perception ne peut s'envisager que comme une boucle sensori-active, mais qu'il y a de plus une interdépendance très forte entre les canaux sensoriels.

Yvette Hatwell met par exemple en évidence l'interdépendance entre la perception de l'espace par le toucher et par la vision et montre par ailleurs que les indices kinesthésiques de localisation spatiale sont traités par le système perceptuel dans un code à la fois kinesthésique et visuel. On peut citer nombre de situations dans lesquelles les modalités perceptives collaborent étroitement pour constituer une perception non seulement multi- mais supra- (ou extra-) modale. Mieux, dans de nombreux cas également, que l'on range dans la rubrique des "paradoxes de la perception" ce qui se passe sur une modalité peut totalement modifier ou transformer ce qu'une autre comprend. L'exemple le plus connu est l'effet McGurk : on donne à entendre à un sujet le son «gaga » tandis qu'on lui fait voir en synchronisme le visage (filmé) d'une personne dont les lèvres prononcent «dada.» Ce que perçoit en fin de compte le sujet n'est ni l'un ni l'autre de ces deux événements, mais «baba.» On peut citer d'autres cas où ce qui se passe sur le plan visuel fait interpréter faussement ce qui se passe sur le plan tactile³, d'autres encore où ce qui se passe sur le plan tactile fait comprendre quelque chose qui était très indistinct sur le plan acoustique ou inversement, etc. On pourrait multiplier les exemples à l'infini.

³ Si l'on donne à manipuler à un sujet un stick doté d'un capteur de force et à voir un spot lumineux se déplaçant proportionnellement aux efforts digitaux appliqués au stick, par une simple modification de la loi du déplacement du spot en fonction de l'effort capté, on peut induire chez le sujet, et ceci de manière systématique, la conclusion de sa part que la résistance du stick à son geste a changé (bien évidemment celle-ci est objectivement constante et liée à des caractéristiques purement mécaniques invariantes, mais on n'en a pas informé le sujet). Ceci est une expérience non publiée, réalisée à l'ACROE.

Le phénomène de perception n'est pas une simple composition des stimuli qui proviennent de la périphérie pour constituer par sommation l'image mentale :

- d'une part, le processus de perception est lui-même globalement une boucle entre les images mentales et les entités du monde extérieur. Le traitement se fait autant dans le sens "bottom-up" que dans le sens "top-down", pour prendre ces termes anglo-saxons (Lindsay et Norman, 1980), ce qui veut dire que le système perceptuel ne procède pas exclusivement en traitant les informations qui viennent de la périphérie (bottom-up), mais formule en permanence des hypothèses sur ce qui peut survenir à l'extérieur en puisant dans un réservoir d'images mentales et de situations apprises antérieurement et se contente dans bien des cas de vérifier que l'hypothèse faite à un moment donné n'est pas contredite par les indices fournis par les sens (top-down).

- d'autre part, s'il y a de toute évidence des images mentales qui sont spécifiquement visuelles ou spécifiquement acoustiques, etc. elles ne constituent qu'une faible partie de l'univers des représentations mentales dont l'existence est pour une large part au moins pluri-modale si ce n'est complètement supra-modale.

En psychologie expérimentale, sur ce deuxième point, deux hypothèses s'affrontent encore : l'hypothèse *centraliste* selon laquelle les données perceptives reçoivent toutes un même traitement, quelle que soit la modalité sensorielle d'où elles proviennent, et l'hypothèse *périphérique* selon laquelle le mode de traitement est spécifique à chaque modalité et dépend des caractères originaux de la réception et de l'organisation corticale de chacune d'elles. Yvette Harwell montre que les divers résultats obtenus en comparant les performances tactiles d'aveugles-nés aux performances visuelles correspondantes des voyants plaident en faveur de l'hypothèse centraliste. Mais au delà de toutes les données expérimentales, un argument simple permet d'affirmer que la perception est non-seulement sensori-motrice, mais fondamentalement multi-sensorielle, c'est que son rôle, pour le vivant, n'est pas la contemplation visuelle, auditive ou tactile, mais l'information, et parfois aussi efficacement que l'urgence des situations concrètes l'exigent, de ce qui survient à l'extérieur. Ce qui survient à l'extérieur se décrit en premier lieu en dehors des modalités sensorielles : ce peut-être un danger physique ou l'éventualité d'un inconfort, ce peut être l'avertissement de la présence d'une proie ou d'un prédateur, ce peut être une relation de collaboration, de conflit, affective ou utilitaire avec un congénère, etc. Certes, l'homme civilisé et cultivé est paré et codé pour toutes ces situations, mais disons que c'est peut-être ce qui l'autorise alors à s'intéresser gratuitement à ses canaux sensoriels et à les étudier séparément sans danger.

1.2. La main organe d'action : le geste ergotique

De la même façon que la perception tactilo-proprio-kinesthésique engage en réalité le corps dans son entier, la main qui agit, transforme, transporte n'est que l'extrémité du système moteur complet qu'est le corps humain, avec sa charpente articulée et son ensemble de muscles. Une étude récente, menée par Peter Jan Beek (BEEK, 1989) sur la "dynamique du jongleur" pose que le système de la motricité humaine, qui gère quelques 792 muscles, 110 articulations et 100 degrés de liberté est pour le moins un *système complexe*. Ce système est capable d'assembler un grand nombre de muscles/articulations en unités fonctionnelles, ces unités peuvent être constituées de diverses manières pour la même fonction et ces processus d'assemblage sont stables et reproductibles. Mais laissons l'étude de la motricité en soi à ses spécialistes (cf. quelques éléments bibliographiques à la fin du texte : BEAUBATON, 1980, PAILLARD et BEAUBATON, 1980, BEAUBATON, 1983). Ce qui caractérise l'action motrice, la main en train de travailler, c'est qu'elle est en prise directe avec la matière, qu'elle peut la modeler, la transporter, l'usiner, la briser.... Elle lui communique non pas de l'information, mais de l'énergie, elle applique des forces, des déformations et des déplacements aux objets et ces derniers lui résistent de diverses manières, lui renvoyant parfois une part de cette énergie.

De la même façon également, une sorte de fovea effectrice se présente au niveau de la main. Bien qu'aucun mouvement de celle-ci ne puisse être absolument dissocié de ceux de la totalité du corps, il y a une concentration, dans l'action manuelle, au niveau des 23 degrés de liberté qu'elle détient à elle seule.

Enfin, à un niveau purement anthropologique, on peut évoquer le considérable progrès que représente, dans l'évolution, l'apparition de la faculté d'"opposition" du pouce. On peut aussi se livrer à une typologie des gestes : gestes préhensiles et non préhensiles, etc. Mais cela ne suffirait pas à rendre à la juste mesure l'immensité des ressources qui permettent à l'homme, du travail de force à la manufacture des objets les plus fins, de structurer à l'extrême son environnement matériel.

1.3. La main organe d'expression : le geste sémiotique

J'utilise ce terme pour désigner l'ensemble des comportements gestuels qui ont pour fonction non plus de connaître ou d'agir, mais de faire connaître : les gestes qui produisent un message informationnel à destination de l'environnement (vivant naturellement). Là encore, ce qui est remarquable c'est la variété des situations ; et là aussi, la main et les membres supérieurs ne sont que les extrémités les plus actives de l'organe d'expression global qu'est le corps dans son entier.

Le geste qui accompagne la parole

Il y a une gestique sémiotique toute naturelle et spontanée, plus ou moins développée selon les cultures mais particulièrement irrépressible dans certaines ; c'est le geste qui accompagne la parole. Contrairement à ce que l'on croit habituellement, ce type de gestes n'est pas seulement un ajout à la parole destiné à l'expression des états affectifs ou à "colorer" l'articulation verbale. On pensait d'ailleurs que leurs processus de production étaient totalement indépendants de ceux de la parole, or il apparaît (McNEIL, 1985, FEYEREISEN, 1987, McNEIL, 1987) que les gestes partagent avec la parole certaines étapes de traitement et que, en conséquence, ils font partie de la même structure communicative ou expressive.

McNeil (op. cit) accorde une certaine importance aux *illustrators* (selon sa terminologie), qui peuvent être iconiques, métaphoriques ou rythmiques, par opposition aux *emblems*, notion qui dans d'autres thèses sont ces actes non-verbaux qui ont une traduction verbale directe (EKMAN & FRIESEN, 1972). La divergence est relatée dans un débat apparemment animé entre McNeil et Feyereisen (FEYEREISEN, 1987, McNEIL, 1987). Cette gestique est donc un mode d'expression à part entière, même si elle est très imbriquée avec la parole. Elle peut faire l'objet d'une étude en soi pour mettre en évidence sinon un langage, au moins une typologie comme celle proposée par Ekman et Friesen (EKMAN et FRIESEN, 1972) qui distinguent alors parmi les *illustrators* :

- les *batons* : mouvements qui accentuent ou donnent de l'importance à un mot particulier ou une phrase ;
- les *ideographs* : mouvements qui esquissent le chemin ou la direction de la pensée ;
- les mouvements *déictiques* : qui indiquent un objet, un lieu ou un événement ;
- les mouvements *spatiaux* : qui décrivent une relation spatiale ;
- les mouvements *rythmiques* : qui décrivent le rythme ou le repos d'un événement ;
- les *kinétographes* : mouvements qui décrivent une action du corps ou quelque action non humaine ;
- les *pictographes* : qui dessinent une figure ou la forme du référent ;
- et les mouvements *emblématiques* qui sont des ensembles utilisés pour illustrer un propos verbal, répétant ou se substituant à un mot ou une phrase.

On peut tirer de ce premier point que les attitudes qui considèrent le geste, contrairement à la parole, comme inapte à toute structuration dans un langage sont peut-être un peu rapides.

Le geste langage utilitaire

Il est abusif de parler de langage à propos du "langage des sourds-muets", car plutôt qu'un langage, c'est un alphabet ou plus précisément un intermédiaire entre l'alphabet et le vocabulaire. Mais ici aussi il y a la démonstration de la capacité du canal gestuel à émettre de l'information codée et structurée.

La gestique du chef d'orchestre est du même ordre en ce qu'elle s'adresse à la vision, l'audition étant mobilisée pour une autre tâche. On ne peut pas non plus la comparer à un langage comme le langage verbal avec un vaste vocabulaire et une syntaxe complexe prédéfinis, conventionnels et stables, le sens émergeant de la combinaison d'éléments qui, pris isolément ont une morphologie arbitraire. Toutefois, le geste du chef d'orchestre a une fonction sémantique (FANTAPIE, 1988). Il comprend une symbolique avec ses règles, ses idiomatismes et ses particularismes. Cette symbolique doit être à même d'exprimer la sémantique musicale et d'ouvrir le dialogue avec les musiciens. La gestique du chef d'orchestre s'appuie sur certaines bases permanentes, mais elle est créée aussi à partir d'une profonde analyse et intériorisation de l'œuvre. Elle est enfin apprise à l'orchestre, à la fois par mode convention et par mode expression/répétition. La gestique du chef d'orchestre est une technique en soi, d'autant plus expressive qu'elle est codifiée et maîtrisée. Il y a des éléments gestuels universels adaptables à toutes les situations, distribués dans un *espace gestuel* délimité dans l'espace et comportant une répartition des tâches entre les bras. Les directions (significations) principales sont la naissance, vie et mort du son, le tempo et la pulsation, le phrasé et le mode d'enchaînement des sons (conduite), les nuances.

Le geste langage esthétique

Une bonne illustration de la richesse expressive du geste est sa manifestation sous diverses formes esthétiques comme dans les ombres chinoises, les arts chorégraphiques orientaux et l'art chorégraphique tout court.

Toutes ces formes d'expression gestuelles ont en commun, qu'elles soient utilitaires ou esthétiques, de s'adresser à l'œil, d'une part, de s'exécuter sans intermédiaire matériel d'autre part. Dès lors que l'on prend en considération la présence d'un tel intermédiaire, la typologie du *geste sémiotique* s'étend.

L'écriture et le dessin : le "Geste Graphique"

L'une utilitaire, l'autre esthétique, l'écriture et le dessin illustrent deux pôles du geste graphique où la manière d'émettre l'information est complètement différente, mais où s'établit un rapport déterminant avec un objet intermédiaire (le crayon, la plume, le pinceau, le marqueur d'une manière générale).

Dans l'écriture, les gestes sont catégorisés et doivent conduire à la réalisation d'objets prédéfinis conventionnellement (les lettres). La quantité de ces objets, donc des différents gestes qui les réalisent donne une mesure de la quantité d'information que le canal gestuel peut transmettre. Toutefois, l'"espace gestuel", c'est-à-dire l'espace physique, mais

également les variétés de dynamiques et de trajectoires caractéristiques de l'écriture, est très contraint ; il n'est pas sûr que le geste d'écriture soit optimal en terme de débit d'information.

Dans le geste graphique esthétique - le dessin - l'espace gestuel (au sens défini plus haut) est beaucoup plus ouvert. Les variétés de formes, mais aussi de manière de marquer sont beaucoup plus grandes. De ce fait, et parce que les langages esthétiques sont par essence ouverts et évolutifs, il n'y a pas les contraintes du geste d'écriture, ce qui permet à une circonstance importante de se manifester : manipuler un crayon ne se fait pas de la même façon que manipuler une plume, un pinceau ou un rouleau. La nature du marqueur a une influence sur l'exécution du geste et par conséquent sur la nature de l'information émise, ce qui n'est évidemment pas le cas dans l'écriture, surtout depuis le développement des stylos à billes et autres feutres qui ont évacué les belles écritures cursives à pleins et à déliés de nos grands-parents.

Le geste sur clavier

Le geste, dans toutes les situations précédentes, avec ou sans intermédiaire matériel, a une finalité visuelle. Avec les claviers, une circonstance doublement nouvelle apparaît : tout d'abord, dans son principe, le clavier "catégorise", c'est-à-dire comme dans le geste d'écriture, il transforme les gestes en un nombre fini de résultats prédéfinis, éliminant par essence les petits écarts d'exécution. Mais ensuite, le clavier installe la première situation où la forme finale du message créé par le geste n'est pas nécessairement visuelle. On oublie en général que le premier clavier fut pour la musique. C'est Ktésibios, mécanicien d'Alexandrie qui l'inventa, en 270 avant J.C. pour commander les sons d'un nouvel instrument : l'orgue. Le mouvement de la touche-glissière de l'ancêtre du clavier est horizontal ; le va-et-vient ouvre et ferme le tuyau correspondant ; un ressort replace la touche dans sa position de départ. Sur ce clavier primitif, le mouvement horizontal de la touche-glissière nécessite une manipulation à pleine main...(HAURY, 1988).

Ainsi, le premier clavier transforme-t-il le geste en son. Depuis cette forme primitive, le clavier s'est naturellement beaucoup perfectionné, donnant lieu à diverses possibilités de production et de modulation du son des divers instruments : orgue, tympanon, clavecin, puis le piano tel qu'on le connaît aujourd'hui. Résultat du perfectionnement de l'invention du facteur florentin Christofori en 1709, le clavier du *piano-forte* associe deux formes de comportement gestuel : par le principe des touches, il catégorise, comme on l'a déjà dit plus haut, mais à l'intérieur de cette catégorisation, c'est-à-dire pour chaque touche il permet une très grande variété de nuances (non catégorisées). Par un principe extrêmement perfectionné, la touche du piano autorise et prend en compte, malgré la petitesse de la course mécanique du doigt, une très grande variété de formes et d'intensités gestuelles toutes converties en variétés auditives différenciables.

Le clavier de machine à écrire, inventé beaucoup plus tard, est un retour à la finalité visuelle du geste catégorisé.

Le geste de commande avec les dispositifs électroniques

Evidemment, avec l'évolution technologique et l'apparition de toutes sortes de possibilités de transformer les forces, déplacements, vitesses, accélérations en signaux

électriques, le geste, s'appliquant à des objets de toutes les formes possibles, a pu finalement s'appliquer sans prédétermination a priori de la modalité destinataire. Au lieu de convertir le geste en signes graphiques ou en phénomènes d'une autre modalité, les transducteurs gestuels divers (les syntaxeurs, comme on dit aussi) "captent" le geste, c'est-à-dire prennent en charge pour la première fois l'aspect informationnel intrinsèque du geste. Ils transforment les manifestations gestuelles en signaux électriques portant l'information gestuelle, dont on dispose alors pour tous les usages voulus.

Mais alors, on voit que le geste *sémiotique* ne peut pas véritablement s'étudier en soi. Beaucoup plus encore que dans le geste graphique, il est dépendant du dispositif auquel il s'applique. Il faut faire une typologie non plus seulement du geste, mais du geste qui s'applique à un objet, c'est-à-dire des couples geste-objet, ce qui donnerait des éléments de repérages tels que ceux de la classification de J.P. Gaillard (GAILLARD, 1990) en trois types de "syntaxeurs" : les poignées, les pantins et les exosquelettes, et en plusieurs types de commandes : les commandes en tout ou rien et les commandes proportionnelles en position ou en effort, etc.

Le geste Instrumental

Enfin, parmi tous les gestes producteurs d'information, il en est un particulièrement remarquable, c'est le geste instrumental musical. On l'a déjà évoqué à propos du geste sur clavier, mais très partiellement. Si l'on considère l'ensemble des instruments de musique, ou au moins les grandes catégories qui peuvent se définir précisément en fonction de la nature de ce qui vibre et de ce qui permet à la main de le faire vibrer, on trouve une immense variété de façons de convertir le geste en information sonore. Ainsi, le geste instrumental musical est-il un geste sémiotique par excellence, qui s'adresse à l'ouïe et dont on peut juger de la capacité informationnelle en considérant la richesse de ce que l'ouïe peut recevoir à travers la musique.

Mais ce que le geste instrumental musical a de singulier, c'est qu'il a tous les attributs (ou presque) du geste ergotique : la main est en contact direct avec des objets physiques (touches, archet, mailloches, etc.) qu'elle manipule, déplace, percute, auxquels elle communique une certaine énergie et desquels elle en reçoit en réaction. Les instruments de musique, prolongements de la main, ont tous les attributs de l'outil, la différence fondamentale cependant est que ce qu'ils permettent de fabriquer (un événement sonore) n'est pas un objet matériel et que la finalité du geste instrumental est informationnelle.

Le geste instrumental est varié : on pince, frotte ou frappe dans le cas des *cordophones* (instruments à cordes), on secoue, racle, entrechoque dans celui des *idiophones* (instruments où les corps vibrants ne produisent pas de sons à hauteur déterminée), etc. Plus fondamentalement, on peut dresser une typologie précise à l'intérieur de grandes catégories telles que *exciter* (fournir de l'énergie à l'instrument qui la convertit en énergie sonore), *contrôler* (ajuster ou modifier un paramètre physique telle que par exemple la longueur d'une corde), *registrer* (modifier ou transformer la structure même de l'instrument, comme dans l'orgue lorsque l'on change la série de tuyaux sollicités). Les catégories de cette typologie sont, usinage en moins, les mêmes que celles du geste ergotique. Elles se fondent sur la présence d'un objet matériel intermédiaire, si bien qu'une typologie du geste instrumental ne peut, en toute rigueur qu'être une typologie *geste-instrument*. Cet objet intermédiaire, foyer de l'interaction, structure l'activité dans une articulation *permanence/variation* fondamentale : la matérialité de l'objet fonde une permanence de base, cohérente avec les propriétés naturelles de l'univers physique ; la permanence de l'objet constitue un substrat pour la transformation

geste->phénomène sonore et, sur cette permanence, deux formes variationnelles hiérarchisées peuvent s'installer :

- variation quantitative (ou paramétrique), par essence continue, analogique, délimitant une classe bien définie de gestes en relation avec des traits morphologiques spécifiques du phénomène sonore,
- variation qualitative (ou structurelle), par essence discontinue, catégorielle, délimitant à son tour une classe de gestes et en conséquence d'aspects sonores, différente de la précédente.

Aux dimensions typologiques précédentes, il faut ajouter celle de la *sélection* : typique du geste sur clavier, elle pourrait se ramener au geste *déictique* si précisément il ne s'agissait pas fondamentalement d'autre chose. Manipuler un clavier ne se résume absolument pas, comme certains avis superficiels le croient parfois, en une simple désignation des touches. Les touches sont sélectionnées *et* actionnées, ce qui n'est pas du tout la même chose.

Enfin, à l'intention de ceux qui avancent que le geste ne peut se plier à aucun langage, je me permettrais de répondre ici que si le geste instrumental n'est pas stabilisé dans un langage standardisé, ce n'est pas par pauvreté, mais c'est pour ne pas perdre sa richesse et son ouverture. Et pour qui ne l'aurait pas senti au travers de ce qui précède, il y a une très forte proximité entre les principales entités constitutives du geste et celles du langage verbal : la partie purement déictique du geste de sélection a la même fonction que le *nom*, le geste d'excitation, actif, la même que le *verbe*, et le geste de modification la même que l'*adjectif*, pour limiter la comparaison à ces quelques éléments.

Quoi qu'il en soit, ce qui est particulièrement essentiel à propos du geste instrumental, c'est que c'est précisément parce qu'il est ergotique qu'il est aussi efficacement informationnel. C'est parce qu'il est matériel, fondé sur un échange énergétique contraint par des spécificités dynamiques, morphologiques, articulatoires qu'il atteint de telles performances expressives. C'est parce que l'archet est archet et non clavier, que le phrasé des instruments à cordes frottées a quelque chose de différent à dire que celui du piano. C'est parce que la mailloche rebondit sur la membrane tendue que la dynamique de la percussion a sa charge émotionnelle particulière.

Certes, le geste instrumental suppose la mise en œuvre de dispositifs très particuliers : l'évolution des archets, des mailloches diverses de percussion, etc. est une histoire aussi longue que celle du clavier. Certes, il suppose également une compétence non banale de l'"opérateur", mais la virtuosité dans le geste instrumental est une situation d'exception qui est aussi (c'est-à-dire pas plus) nécessaire au geste instrumental que (ne) l'est la virtuosité stylistique à l'expression écrite...

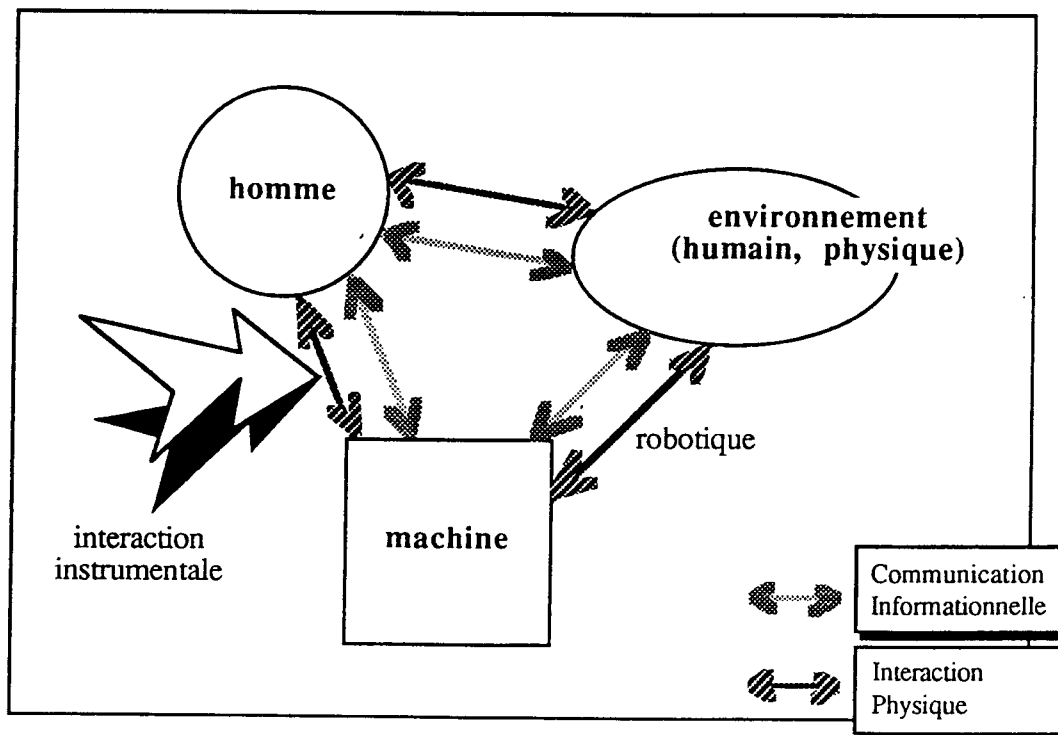
2. L'Interaction Homme/Machine/Environnement

Venons-en à la machine. Le titre de ce chapitre aurait pu être "canal gestuel et machine", ou "la communication gestuelle avec la machine". Que l'on me pardonne ce qui peut être pris pour d'interminables préalables, mais c'est trop souvent à force de foncer dans les directions toutes tracées que l'on se retrouve au bout du chemin en ayant raté l'essentiel.

On parle indistinctement de *Communication* et d'*Interaction* Homme/Machine. J'ai déjà souligné le caractère anthropomorphique du terme *Communication* employé à propos de la relation avec les machines. Il n'est de sens à la communication qu'entre les hommes et s'il faut introduire la problématique complète des relations entre l'homme et

l'ordinateur, c'est en tenant compte de deux grandes catégories différentes : celle où l'ordinateur est un intermédiaire entre l'homme et l'homme, où le terme de communication est légitime et celle où l'ordinateur est un intermédiaire entre l'homme et le monde physique, où il ne l'est pas (on pourrait considérer une troisième catégorie dans laquelle l'ordinateur est un intermédiaire entre une partie du monde physique et une autre, mais c'est un cas limite car l'homme n'en est jamais complètement absent).

J'ai ensuite considéré que l'anthropomorphisme était tolérable dans la deuxième situation en tant que métaphore utile et positive. Je voudrais maintenant montrer qu'en fait, il est incomplet. En effet, on donne en général de la relation entre l'homme, la machine et l'environnement, une représentation synthétique sous la forme d'un schéma triangulaire qui fait apparaître une liaison bidirectionnelle entre chacun des sommets du triangle (voir la figure page suivante). La machine s'introduit entre l'homme et son environnement, vivant et physique, mais elle introduit également, et c'est cela la pertinence de ce schéma triangulaire, des relations spécifiques machine-environnement, et ne supprime pas complètement les relations directes homme-environnement. Ce que l'on ne pense pas de faire valoir alors, c'est qu'en chapeau général de la typologie des relations, qu'il s'agisse de celle de l'homme avec l'environnement, de la machine avec l'environnement, *et même de l'homme avec la machine*, il y a un partage en deux catégories fondamentales qui, bien que parfois très imbriquées, sont distinctes et complémentaires. Pour les trois interactions, sans exception, on peut envisager d'une part des relations communicationnelles (épistémiques et sémiotiques) d'autre part des relations ergotiques. Les premières ont lieu entre l'homme et la partie humaine de l'environnement, entre l'homme et l'ordinateur dans les différentes situations que l'on a déjà évoquées et entre l'ordinateur et la partie humaine de l'environnement. Mais les secondes ont lieu aussi dans les trois cas : de l'homme à son environnement physique lorsqu'il y intervient matériellement, de l'ordinateur à l'environnement physique dans tout ce qui relève de la *robotique*. On oublie alors en général que la relation entre l'homme et l'ordinateur peut être elle aussi ergotique. C'est cette relation qu'il semble intéressant de considérer. On peut la désigner par *relation instrumentale*.



Les bras articulés et autres dispositifs à moteurs contrôlés par l'ordinateur de la robotique sont des dispositifs qui "informent" l'énergie motrice et qui restituent à ce titre la substance du phénomène ergotique dans le sens homme -> environnement : le mouvement et l'effort structurés en actions. Le principe technologique qui est employé dans ce sens est tout à fait applicable dans l'autre, pour donner à l'opérateur une restitution de la réaction du monde physique à ses gestes matériels ; il suffit de disposer des moteurs contrôlés par l'ordinateur lui-même, dans les organes de commande manipulés par l'opérateur.

En conséquence, le vrai sujet est celui non seulement de la *Communication* entre l'homme et la machine, mais de *l'Interaction entre l'homme, la machine et leurs environnements*. La relation entre l'homme et la machine se décline alors selon trois aspects : dans le premier, la forme et la nature de l'interaction auront intérêt à se décrire en référence aux archétypes de la communication humaine naturelle, dans le deuxième, elles prendront référence dans les relations, non moins naturelles, de l'homme et de son environnement physique. Ces deux formes peuvent s'extrapoler et s'imbriquer intimement, mais il y a un troisième aspect : celui où la machine introduit ses propres spécificités, irréductibles aux deux autres catégories...

3. L'introduction du Geste dans l'Interaction Homme/Machine

3.1. Interfaces, Capteurs, Effecteurs, Transducteurs, ...

La machine informatique est un univers de phénomènes qui n'interagissent pas spontanément avec l'environnement ou avec l'homme ; tout n'y est qu'*information*, or l'information est un concept, elle n'"existe pas" dans la nature. Dans la nature il y a des phénomènes physiques et parmi ceux-ci certains auxquels l'homme est sensible ou qu'il peut produire ; et la question est que les phénomènes physiques qui matérialisent le concept d'information dans l'ordinateur sont précisément étrangers aux sens et aux actes humains ; de même qu'ils sont le plus possible isolés des phénomènes de l'environnement. Il ne peut alors y avoir interaction, en particulier entre l'homme et la machine informatique qu'à la condition d'une part que l'information qu'elle destine à l'homme prenne corps dans un phénomène susceptible de l'atteindre, d'autre part que les actes de l'homme puissent être convertis en information au sens de la machine.

Dans la technologie informatique, il y a alors des organes spécifiques dont la charge est de réaliser cette relation entre les univers phénoménologiques humains et/ou physiques spécifiques et l'univers informationnel électronique. On les appelle d'une manière générale les *interfaces*. Il y a ainsi un problème général d'*interface* que l'on peut définir de la manière suivante : caractériser les aspects de la communication du point de vue des conditions et des fonctions côté homme, définir et construire les organes matériels et les fonctions logicielles appropriées. Pour introduire correctement le problème spécifique au canal gestuel, il convient ici d'établir certaines distinctions.

Compte tenu de ce que l'on envisage la plupart du temps seulement deux formes restreintes d'interaction : celle de l'homme et de la machine d'une part, que l'on considère de type exclusivement informationnel, et celle de la machine avec l'environnement physique que l'on considère comme informationnelle de l'environnement vers la machine et ergotique de la machine vers l'environnement (c'est la robotique), on a été amené, finalement à ne définir qu'une liste incomplète des types d'interfaces.

Dans la typologie des interfaces il faut bien faire intervenir deux composantes distinctes :

- la composante purement informationnelle, dans laquelle il s'agit dans un sens de convertir un phénomène pour n'en retenir que l'information (à énergie minimale) en un phénomène électronique recevable par la machine, dans l'autre de produire un phénomène sensible à partir de cette information ;
- la composante ergotique, dans laquelle il s'agit, comme on l'a déjà dit plus haut d'"informer l'énergie", c'est-à-dire de passer d'un phénomène purement informationnel puisqu'il siège dans la machine informatique, à un phénomène physique énergétique structuré.

Mais la plupart du temps, pour cette deuxième composante, on s'en arrête là, c'est-à-dire au sens machine->environnement. Ce qu'il faut introduire, au niveau des principes, et donc à la base de la typologie des interfaces, c'est le symétrique de la fonction qui vient d'être décrite. C'est en particulier là une base essentielle pour le canal gestuel, si l'on veut que dans sa relation avec la machine, l'homme puisse exploiter cette faculté importante dont il dispose dans sa relation avec l'environnement : le geste ergotique.

Interface gestuelle

En conséquence, pour le canal gestuel, le concept d'interface est non seulement tout aussi fondamental que pour tous les autres canaux de communication et d'interaction, mais il doit a priori prendre en compte, différemment selon les applications bien évidemment, l'ensemble des fonctions suivantes :

- capture des phénomènes gestuels et extraction de l'information caractéristique non seulement de leur forme, mais de leur comportement ergotique, c'est-à-dire de leur énergie (ce qui suppose la prise en considération des deux variables duales de la mécanique dans la communication gestuelle) ;
- production des phénomènes correspondant à une réaction physique mécanique en réponse au geste ergotique, ces phénomènes étant par ailleurs les stimuli du sens tactilo-proprio-kinesthésique.

Capteurs et effecteurs, Transducteurs,...

Chez l'être humain, on peut considérer les fonctions analogues de celles qui vont établir la relation entre la périphérie de la machine et ses processus les plus internes. Cependant, il y a une condition très particulière à la technologie, tout du moins dans son état actuel, qui fait que le parcours de la périphérie vers l'intérieur et le retour, dans la machine, se font en passant par une suite de fonctions très catégorisées, matérialisées séparément ou de manière imbriquée, par des organes très spécifiques⁴.

⁴ On peut considérer les mêmes fonctionnalités d'interfaçage chez l'homme, entre la périphérie et le "traitement" par son cerveau. Cependant, il convient d'être prudent et raisonnable : tout d'abord, le concept d'information n'est peut-être pas applicable aussi simplement à l'homme qu'à la machine. Ensuite, anthropomorphie pour anthropomorphie, il se peut bien que la machine se venge un petit peu : il existe à l'égard de l'homme et du cerveau l'exacte contrepartie de l'attitude anthropomorphique à l'égard de la machine, dans la tendance à expliquer la construction et le fonctionnement de l'être humain en y projetant les formes et catégories spécifiques à l'informatique. Certains comme Minsky ou Hofstadter vont jusqu'à considérer que la pensée est un algorithme. Les plus grands pionniers sont parfois plaisants par la fraîcheur de leur naïveté.

Chez l'homme, il y a des récepteurs, c'est-à-dire des structures cellulaires qui sont sensibles à diverses catégories de phénomènes, qui deviennent ce que l'on appelle les phénomènes sensibles. Il y a des impulsions électriques conduites par les nerfs et les neurones, il y a une structure très complexe de

Plus précisément dans le sens homme->machine comme dans l'autre, on rencontre :

- La fonction *capteurs / effecteurs*. On utilise parfois le terme de *transducteur* pour évoquer les organes qui convertissent un phénomène physique donné en un autre phénomène physique en préservant entre les deux une identité (plutôt une correspondance) morphologique. Ce terme vaut donc pour les capteurs et les effecteurs : la bobine mobile du microphone, ou du haut-parleur, la surface photosensible du tube de la caméra, l'enduit luminophore, le faisceau d'électrons du tube cathodique. Pour le canal gestuel, on trouvera des jauges de contrainte, des cristaux piézo-électriques, des moteurs, des ballonnets gonflables, etc. On y reviendra plus loin.

- La fonction de *plongement* du dispositif matériel de communication dans l'espace phénoménologique humain ; cette formulation compliquée adresse un problème simple, mais extrêmement important : il ne suffit pas que les phénomènes soient visibles, audibles, tactiles, etc., il faut qu'ils se développent dans un certain espace. Il s'agit par exemple de l'organisation des points lumineux dans une zone plane (écran) ou dans un espace tridimensionnel (dispositifs de visualisation stéréoscopique), de la diffusion sonore monorale ou stéréophonique, de la captation omnidirectionnelle ou monodirectionnelle de la voix, des espaces et trajectoires de développement du geste etc. Dans le cas du canal gestuel, les capteurs et les effecteurs sont montés sur des dispositifs mécaniques irréductibles qui conditionnent fortement l'espace, les trajectoires, le nombre de degrés de liberté des phénomènes gestuels. Il s'agit de la caractéristique *morphologie* des interfaces gestuelles.

- Enfin, directement associés aux transducteurs, il y a dans le cas général des dispositifs chargés de "conditionner" (ou "déconditionner") les signaux spécifiques à chaque type de transducteur avant leur entrée dans (ou après leur sortie de) l'espace de traitement informatique.

Je vais présenter maintenant quelques unes des principales réalisations et expérimentations qui, de par le monde (informatique), relèvent d'une manière ou d'une autre de la prise en charge du canal gestuel dans l'interaction homme/machine.

3.2. Les systèmes et dispositifs gestuels existants

Les technologies des transducteurs gestuels

Dressons tout d'abord un inventaire succinct des technologies actuellement utilisées pour capter les phénomènes gestuels et produire les "retours physiques". Elles sont en fait très dépendantes de la nature du dispositif et plus précisément de la fonction qu'on entend lui faire remplir. Il y a à la base deux catégories de situations différentes : celles où le geste s'applique à un objet et celles où le geste s'effectue sans contact matériel d'aucune sorte. Il est clair que seule la première peut éventuellement inclure une composante ergotique ; ce qui ne disqualifie pas a priori la seconde puisque, on l'a vu,

traitement de ce que l'on peut bien appeler des signaux, électriques, chimiques, etc, le cerveau. Il y a enfin des organes effecteurs, nos muscles essentiellement. Il ne s'agit cependant en aucune manière d'une chaîne linéaire qui irait, via le cerveau, du bout de nos cellules réceptrices aux fibres de nos muscles mais d'un immense système de boucles imbriquées. Dans l'homme, autre preuve de ce que la machine en est encore très loin, les fonctions de réception et de traitement, de traitement et de mémorisation, de réception et d'émission ou d'action ne sont pas aussi localisées et dissociées que dans la machine ; et ceci n'est sans doute pas sans incidence et sur la performance et sur la signification.

une classe intéressante de gestes (sémiotiques) existe sous les catégories du *geste accompagnant la parole* ou du *geste du chef d'orchestre*, etc. Il faut seulement ne pas compter, avec des dispositifs sans contact, pouvoir exploiter les mêmes propriétés et performances gestuelles qu'avec les dispositifs à contact. Le choix de l'une ou l'autre de ces deux situations conditionne explicitement et complètement le type de geste qui pourra être mis en œuvre.

Dispositifs sans contact

Bien évidemment il ne peut être question ici que de *capteurs* car on n'imagine pas par quelle magie une force ou une énergie pourrait être communiquée à la main en l'absence de tout intermédiaire matériel.

Les techniques utilisées dans cette situation sont des techniques de *télécapage* :

- la saisie optique des configurations et mouvements de la main (ou éventuellement du corps tout entier) par une caméra. L'information vidéo est traitée de différentes manières pour extraire une information significative. Le système d'analyse d'image qui doit donc suivre est en général complexe parce qu'il doit se contenter d'une information bidimensionnelle (ou d'une double information bidimensionnelle si l'on place deux caméras en stéréoscopie) pour "interpréter" un phénomène spatial. De plus, il n'y a aucune isomorphie a priori entre le procédé d'acquisition de l'information visuelle par un balayage vidéo et la structure de l'espace des mouvements gestuels. Ces dispositifs ne permettent d'extraire qu'une information gestuelle réduite et simplifiée, ou alors on entre dans la branche plus fondamentale de la *vision artificielle* mais il ne s'agit plus à proprement parler d'une simple capture du geste.

Une variante de ce principe est sa transposition dans l'espace des ondes ultrasonores. L'opérateur est "éclairé" par des sources d'ultrasons et la caméra est remplacée par des microphones appropriés. La résolution et la précision sont naturellement encore plus faibles.

- Les autres techniques consistent soit à placer un élément capteur, soit un élément émetteur sur les membres mobiles. Ces éléments doivent être très légers sans quoi on sortirait de la catégorie "sans contact". Dans le premier cas, le principe le plus performant est celui du capteur Polhemus⁵. Ce capteur utilise trois petites bobines électromagnétiques dont les axes sont orthogonaux, noyées dans un corps en plastique. Fixées sur le membre mobile, elles fournissent un signal électrique dépendant de leur position dans un champ magnétique créé par un autre dispositif distant. L'information captée n'est que celle du mouvement des points où se situent les capteurs, mais elle est intrinsèquement spatiale ; aucun traitement spécifique n'est nécessaire pour restituer l'espace. Dans le second cas, on trouvera l'usage à nouveau d'une caméra et de LED placés sur les doigts ou de micros ultrasoniques et de sonars placés sur la main. Dans le cas des LED, la bande passante peut être élevée (de l'ordre de 200 Hz), et le mouvement de chaque doigt est capté, mais l'information doit être analysée.

On peut signaler enfin une technique mise en œuvre dès la fin des années 20 par le physicien russe Teremen pour capter le mouvement des mains d'un musicien afin de lui permettre de contrôler l'amplitude et la fréquence de sons engendrés par un système électronique. Le dispositif de Teremen comportait deux antennes propageant un champ électromagnétique haute fréquence à faible puissance ; chaque champ était altéré en fonction de la proximité des mains, ce qui constituait la base d'une détection,

⁵ Nom de la société qui commercialise ce dispositif.

indépendante, du mouvement de chaque main. Ce dispositif était relativement précis, mais il ne donnait de l'espace gestuel que deux variables représentatives. On peut remarquer au passage qu'il exploitait un geste type *chef d'orchestre* à des fins plutôt instrumentale. Mais c'est une autre discussion.

Dispositifs avec contact - Capteurs

Dans le cas des dispositifs avec contact, tous les systèmes, quelle que soit leur nature et leur sophistication peuvent se ranger dans l'une ou dans l'autre des deux grandes catégories suivantes : les *dispositifs à immersion* et les *dispositifs à vis-à-vis*. Il s'agit de deux "philosophies" différentes. Pour des raisons matérielles incontournables elles ne peuvent pas avoir les mêmes technologies de base sur tous les aspects et par conséquent les mêmes performances. Mais surtout, elles correspondent à des visions très différentes du rapport de l'homme avec la machine. Je reviendrai plus loin sur ce point.

Dispositifs à immersion

Ici on "instrumente" l'opérateur, c'est-à-dire qu'on le considère en soi, comme ses gestes d'ailleurs, et on prend le parti de mesurer le plus précisément possible toutes les variables significatives de ses comportements gestuels. C'est en quelque sorte de la proprioception extravertie. Les solutions techniques à ce problème ne sont pas simples car il y a une contradiction de principe, entre l'exhaustivité et la précision de la mesure d'une part, et la neutralité du dispositif par rapport à ce qu'il mesure d'autre part. En d'autres termes plus on veut une mesure précise et exhaustive, plus le support matériel des capteurs est lourd et encombrant, ce qui fait que les gestes captés sont ceux d'un opérateur handicapé. Il n'y a pas de solution idéale à ce problème dans l'état actuel de la technologie.

Ce problème s'aggrave lorsqu'il s'agit non plus des capteurs, mais des effecteurs. On y reviendra plus loin.

Les dispositifs à immersion sont d'une manière générale constitués d'une structure matérielle sur laquelle sont montés les transducteurs. On distingue principalement les gants ou combinaisons de données d'un côté (structures support continues), les exosquelettes de l'autre (structures discrètes). La précision de capture ne dépend pas beaucoup du type de structure. Elle est déterminée par celle des capteurs et par leur nombre dans les deux cas. Du gant à la combinaison, il n'y a qu'une différence de degré d'immersion, différence que l'on retrouve entre les exosquelettes qui ne prennent que la main et ceux qui prennent le corps dans son entier.

Dans les technologies de captage du mouvement on trouve essentiellement la détection de déformation de fibres optiques, de jauges de contraintes, le déplacement de curseurs sur pistes potentiométriques. On trouve aussi des capteurs à effet Hall et à effet inductif.

Dispositifs à vis-à-vis

Dans ce cas on envisage par principe la situation comme une interaction : l'opérateur agit sur un objet, lequel fait explicitement partie de la scène, et ce que l'on mesure n'est pas le comportement gestuel intrinsèque mais les variables représentatives de l'interaction entre l'opérateur et l'objet. La présence explicite de l'objet impose alors certaines

contraintes telles que la caractérisation et la prise en considération de ses aspects propres : forme, degrés de déformabilité ou de liberté, trajectoires, articulations, inerties, frottements etc. Mais en contrepartie, la précision et les performances dynamiques peuvent atteindre des niveaux véritablement opératoires.

Les dispositifs à vis-à-vis sont très variés, cela va du bouton rotatif au clavier multi-morphologies en passant par les "joysticks", poignées, bras-mâtres (maxi et mini). La technologie des capteurs est qualitativement la même que pour les dispositifs à immersion.

Dispositifs avec contact - Effecteurs

C'est à ce niveau qu'une discrimination radicale s'opère entre les dispositifs à immersion et les dispositifs à vis-à-vis. Le problème posé aux premiers à propos des capteurs est décuplé lorsqu'il s'agit des effecteurs, ce qui se conçoit aisément dans la mesure où il faut cette fois être non seulement informationnel, mais énergétique. Pour ne pas entraver les gestes de l'opérateur, la présence mécanique propre du dispositif doit être minimum ; c'est ceci qui impose une limite intrinsèque à l'installation du retour physique de l'ordinateur vers l'opérateur. Ce retour peut être de plusieurs ordres différents : **retour simplement tactile**, s'adressant à ce que l'on appelle la *perception tactile cutanée simple*, ou **retour d'effort** s'adressant à la *perception tactilo-proprio-kinesthésique* dont on a vu la richesse et l'importance. Les forces et les déplacements mis en jeu dans le second cas sont sans commune mesure avec ceux du premier : la perception d'un obstacle dur et de son immobilité ou de son déplacement spécifique peut exiger des forces (impulsionnelles) de plusieurs centaines de Kgf tandis qu'il s'agit de grammes dans le cas de la perception d'un état de surface ou de la simple information du fait qu'il y a contact entre la main et cette surface.

Plus généralement, la perception tactilo-proprio-kinesthésique se constitue à partir de phénomènes mécaniques d'une certaine qualité dont les caractéristiques importantes sont : la puissance (que nous avons déjà évoquée), la forte bande passante (sans être évidemment comparable à la bande audible, les forces ou déplacements en jeu peuvent atteindre des fréquences voisines de 1 KHz), le grand nombre de degrés de liberté ou de déformabilité des objets (la main dispose, rappelons-le, de 23 degrés de liberté).

De fait, les performances des effecteurs exigées pour un retour d'effort qui ait un sens au plan communicationnel, sont incompatibles, du moins en l'état actuel de la technologie, avec le principe des dispositifs à immersion. Ces derniers permettent tout au plus une approche du *retour tactile*, ce qui n'a rien à voir.

Les technologies du retour tactile sont actuellement les suivantes : ruban piezo-électrique, alliages à mémoire de forme, bobines vibrantes de type haut-parleur, ballonnets gonflables,... Celles du retour d'effort sont essentiellement basées sur les moteurs électromagnétiques et les moteurs hydrauliques.

Je vais présenter maintenant quelques unes des expérimentations ou des réalisations qui introduisent d'une manière ou d'une autre le canal gestuel dans la relation avec la machine, en m'attachant surtout à faire ressortir la démarche spécifique qu'elles sous-tendent.

Dispositifs et systèmes

Bras-Mâtres et Bras-Esclaves - La Télémanipulation

Un précurseur de la communication gestuelle qui ne pose pas la question de l'information mais de la manipulation matérielle est la *télémanipulation* sous ses diverses formes. Les premiers dispositifs de télémanipulation, mis en œuvre pour manipuler à distance des objets dangereux (radioactifs par exemple), étaient de simples dispositifs de transmission mécanique homomorphe (pantographes). Ils ont été rapidement remplacés par des systèmes utilisant l'électricité comme moyen de transmission. Dans la télémanipulation, il n'y a pas à proprement parler de machine au sens informatique du terme, mais seulement deux organes complémentaires : celui qui capte les mouvements de l'opérateur et les transforme en signaux électriques que l'on peut alors transmettre - on l'appelle en général le "bras-maître" - et celui qui, au bout de la chaîne de transmission effectue l'opération inverse, c'est-à-dire transforme le signal électrique en mouvements images des mouvements de l'opérateur - le "bras esclave". Le bras esclave et le bras maître peuvent être d'ailleurs exactement symétriques l'un de l'autre ; les moteurs montés sur l'un et sur l'autre pouvant jouer indifféremment le rôle de restitution de l'effort de l'opérateur sur l'objet manipulé, du côté de l'objet, ou de restitution de la résistance de l'objet aux mouvements de l'opérateur, côté opérateur.

Il peut y avoir une "machine" entre les deux, purement analogique, c'est-à-dire réalisant par exemple, par amplification/atténuation, un changement d'échelle spatiale et énergétique entre l'espace de l'opérateur et celui de l'objet télémanipulé. Mais l'on ne peut pas véritablement parler de traitement au sens informatique du terme, et, d'autre part, s'il y a une machine, elle n'est qu'un intermédiaire de communication entre l'opérateur et un objet réel. Toutefois, dans la mesure où le signal qui circule entre le maître et l'esclave contient tout ce qu'il faut pour restituer les mouvements de l'un au niveau de l'autre, c'est qu'il renferme une *information* relative à ces mouvements : geste de l'opérateur, réaction de l'objet. De là, il n'y a qu'un pas à franchir pour faire de cette information une entrée et/ou une sortie d'une machine informatique.

Le projet GROPE, Université de Caroline du Nord - Etats Unis

Le projet GROPE, démarré en 1967 par Frederick Brooks au Département d'Informatique de l'Université de Caroline du Nord aux Etats Unis (BATTER, BROOKS, 1971, BROOKS, OUH-YOUNG, BATTER, KILPATRICK, 1990) est probablement le premier à avoir mis en œuvre un dispositif à retour d'effort dans une relation non plus de télémanipulation, mais d'interaction avec un processus informatique. Il a été stimulé par les idées de Sutherland, vers 1965, qui présentait l'ultime dispositif d'interface ("*the ultimate display*") comme une "fenêtre sur le monde virtuel de l'ordinateur", incluant la vue, l'audition et le toucher.

Centré sur une application cible : la recherche de liaisons chimiques entre molécules complexes par simulation et représentation graphique, le projet GROPE a développé en plusieurs étapes divers dispositifs de manipulation, à 2, 3 puis à 6 degrés de liberté, dans le but de restituer à l'opérateur une information tactile (*haptic display*) coordonnée à la visualisation. Dans sa version la plus récente (1988), le dispositif de manipulation est un bras-maître à 6 degrés de liberté, de grandes dimensions (le bras ARM : Argonne Remote Manipulator), muni d'une poignée à l'extrémité. Dans l'une des applications typiques, par l'intermédiaire de ce bras, qui comporte donc moteurs et capteurs, l'opérateur manipule une petite molécule qu'il essaie d'arrimer à une plus grosse (l'arrimage correspond à l'établissement de la liaison chimique) le tout lui étant présenté sur un grand écran. Les molécules sont simulées par un calcul géométrique et un calcul des champs de forces d'interaction. En manipulant la poignée, l'opérateur produit les informations spatiales qui commandent les déplacements de la molécule à arrimer. La simulation calcule alors les 6 composantes de force de l'interaction, qui sont envoyées au

bras-maître. En s'aidant de ses sensations tactiles, l'opérateur peut ainsi découvrir les liaisons chimiques possibles.

L'introduction de la manipulation directe de la molécule virtuelle et du retour *haptique* combiné avec la visualisation est une situation véritablement pertinente : elle est plus qu'une simple aide, elle apporte une forme de compréhension supplémentaire. Par ailleurs, elle permet un véritable processus de découverte : avec le GROPE-III (le dispositif le plus récent, mis au point en 1988), les chimistes retrouvent précisément les positions d'arrimage connues et peuvent découvrir de très bonnes solutions pour les liaisons encore non connues.

Les performances de ce dispositif sont cependant encore relativement limitées : le bras est encombrant, la force maximale ramenée au point de manipulation est de l'ordre de 300g (8oz), et la rapidité de réponse des moteurs ainsi que la fréquence d'échantillonnage des données gestuelles (de 20 à 80 Hz) sont assez faibles. La réalité mécanique propre du bras "pèse" sur ces performances et sa compensation numérique est vite limitée par des instabilités prohibitives pour le fonctionnement fin. Par ailleurs, la lourdeur des calculs de simulation, combinée aux temps de réponse importants de la chaîne électro-magnéto-mécanique du dispositif empêche d'atteindre les temps de boucle nécessaires au canal gestuel (de l'ordre de la ms). La conséquence est que les objets simulés restent relativement "mous" et, plus généralement, que la finesse de restitution des phénomènes pour la perception proprio-tactilo-kinesthésiques reste en-dessous d'un certain seuil nécessaire au réalisme. Un des moyens par lesquels les concepteurs du système ont tenté de pallier à la "mollesse" de l'univers restitué a été d'associer l'émission d'un son percussif au franchissement d'une surface censée être dure, synchrone de l'instant de franchissement. Le résultat est convaincant, mais il est à mettre au compte des vertus de la fusion intermodale et plus précisément de ce que l'on pourrait appeler les *transfuges intermodaux*, ces phénomènes paradoxaux dans lesquels ce qui se passe sur une modalité conduit, en dépit des faits objectifs, à conclure que cela se passe sur une autre. Ici, en l'occurrence, le son percussif induit l'idée de la perception tactile d'une surface dure, mais ce n'est que l'idée et cela ne constitue pas une amélioration de la restitution tactile en elle-même.

Dans le même laboratoire, d'autres dispositifs sont expérimentés, tels ce cyclo-cross à retour d'effort qui permet la visite interactive d'un bâtiment virtuel. Le vélo est fixe, naturellement, mais le cycliste reçoit, à travers un casque vidéo stéréoscopique (Head Mounted Screen), les images d'un décor synthétique qui se modifient en cohérence avec ses mouvements sur le pédalier. Ce dernier est couplé à un transducteur électromécanique qui rend le pédalage plus ou moins facile selon que l'opérateur se trouve dans une montée (virtuelle), un terrain plat ou une descente. Il s'agit là du captage d'un comportement gestuel (corporel) exclusif et entièrement prédéterminé.

VIDEOPLACE, ... *Myron Krueger*

Parmi les pionniers reconnus des "réalités virtuelles", dans lesquelles on prétend volontiers avoir recours au geste, il faut citer Myron Krueger, à l'Université de Wisconsin, qui a mis au point successivement, depuis 1969, un certain nombre de dispositifs (VIDEOPLACE, GLOWFLOW, METAPLAY, PSYCHIC SPACE, MAZE) (KRUEGER, 1977) permettant une "interaction" entre le monde réel des opérateurs et le "monde virtuel" de l'ordinateur, ainsi que des situations étonnantes d'interactions de plusieurs individus distants, dans un même espace virtuel. Ces expériences relèvent plus des "réalités virtuelles" que de la communication gestuelle, mais elles méritent d'être évoquées ici en raison de la démarche particulière qu'elle représentent et aussi parce qu'elles correspondent à une exploitation complètement différente du geste pour

communiquer avec la machine : d'une manière générale, une image globale de l'opérateur est captée par une caméra ; de cette image est ensuite extraite une information simplifiée et significative, le contour par exemple, projetée sur un grand écran, face à l'opérateur. Cette information est alors une entité propre dans un univers visuel virtuel avec lequel elle interagit. L'opérateur développe ses comportements en se servant, comme guide et contrôle, de la vision de son image réduite en interaction avec les objets de l'univers visuel virtuel. Il est facile d'imaginer cette complexification du dispositif dans laquelle on réalise la même situation en deux endroits avec deux opérateurs distants et où l'on restitue à l'un et à l'autre, chacun en leur propre lieu une image commune dans laquelle ils se voient tous les deux. On ne peut pas dire cependant que les opérateurs interagissent ni qu'ils interagissent avec la scène, à moins qu'ils n'admettent de vivre par procuration au travers de leurs images ; car en toute rigueur ce sont surtout elles qui interagissent...

*"Ainsi font..." les marionnettistes à l'Institut de Recherche sur la Perception.
Eindhoven - Pays Bas.*

Le Simple-Simon's GesturePad™ commercialisé par Philips et développé par Gerrisen & al. (GERRISEN, ITEGEM, DAAMEN, 1990) chercheurs de l'*Instituut voor Perceptie Onderzoek* (IPO) de Eindhoven présente des points communs avec l'approche précédente, mais correspond à une exploitation du geste comme langage pour l'homme.

Le dispositif s'inspire de la "tactique des marionnettes" : les marionnettes sont des machines à mimer simples ; leurs postures, mouvements et gestes peuvent connoter, accompagner ou illustrer un dialogue ou une histoire contée. Le contrôle des marionnettes est basé sur un vocabulaire limité de postures élémentaires, qui peut s'enrichir beaucoup selon la dextérité et la sensibilité du marionnettiste. Le système proposé applique cette "tactique" à une marionnette particulière : la main elle-même. Il consiste en un clavier spécial et une souris connectés à l'ordinateur ainsi qu'un logiciel qui affiche à l'écran des postures de mains schématisées prises dans un répertoire prédéfini. La posture qui s'affiche dépend alors de la touche enfoncée, mais d'une façon telle qu'elle corresponde sensiblement à la posture réelle de la main en train d'enfoncer cette touche. Par ailleurs, la souris permet, dans le même temps de positionner spatialement la posture sélectionnée sur l'écran. Deux opérateurs distants, dont les ordinateurs ainsi équipés sont reliés peuvent alors communiquer par le geste. Mais les opérateurs en question peuvent aussi être par exemple un conférencier et son public suivant les explications du premier sur grand écran, lequel représente un certain nombre d'informations et de schémas auxquels se superpose ce dispositif.

Le gant de données VPL

Le gant de données (*dataglove*) de la firme VPL crée par Jaron Lanier et Jean-Jacques Grimaud a fait une entrée pour le moins remarquée aux environs de 1985. Il est, associé au HMS (*Head Mounted Screen*, ou paire d'écrans montée sur les yeux) et au procédé qui consiste à modifier la vision stéréoscopique d'une image synthétisée tridimensionnelle en rapport avec les mouvements de la tête de l'opérateur, à l'origine du courant médiatique des "réalités virtuelles". Ce dispositif mérite absolument le titre de capteur de geste dans la mesure où il permet d'envoyer à l'ordinateur des informations sur la configuration et les déplacements de la main, grâce à un système de fibres optiques et au capteur de position 6D Polhemus installés sur le gant, donc, que l'on enfle. Mais il ne mérite certainement pas toute la publicité qu'on a pu en faire à une époque, si l'on y regarde d'un peu près quant à la finesse et la précision de cette capture. La capture des mouvements par le dispositif Polhemus est relativement précise, mais celle des configurations des doigts par les fibres optiques ne l'est pas du tout. De plus elle présente

de gros défauts de fiabilité et de reproductibilité. Les gestes de la main qui sont ainsi saisis par ce dispositif sont en fait extrêmement rudimentaires, en sorte que le *dataglove* n'est qu'une *souris spatiale*.

La démonstration type de ce dispositif est dans son association au HMS, où l'opérateur se déplace dans un espace virtuel peuplé d'objets (bientôt d'êtres) virtuels et peut les "saisir" en effectuant les mouvements de la main au moment où il entre dans leur proximité. Une difficulté énorme de l'expérience tient au fait que rien n'empêche, au moment d'entrer en contact avec les objets, de les traverser impunément... Cette application est en fait abérante en ce que le dispositif ne permet pas un geste d'un autre type que celui du langage (des sourds muets ou du chef d'orchestre) mais qu'elle prétend l'utiliser pour des opérations de type instrumental.

Le gant CyberGlove, de la société Virtex

Ce gant, conçu par Jim Kramer, fondateur de *Virtex* à Stanford reprend le même principe, mais avec une autre technologie : 22 capteurs à jauge de contrainte. Incomparablement plus performant, ce dispositif permet effectivement d'exploiter la main dans une situation communicationnelle du type "langage des sourds-muets" ou "geste du chef d'orchestre". Une application typique et parfaitement au point consiste, ayant "chaussé" le gant, à effectuer un certain nombre de configurations de la main en demandant à l'ordinateur de les "apprendre" et de les associer à des lettres de l'alphabet. Celui-ci effectue cette tâche grâce à un logiciel type réseau de neurones. Dans une seconde phase, l'opérateur effectue les gestes appris par le logiciel, lequel identifie alors les lettres correspondantes. Le nombre de lettres ainsi identifiables est largement égal au nombre de lettres de l'alphabet, ce qui donne une mesure des capacités discriminantes du système de capture.

A la suite des gants de données viennent les "combinaisons de données" qui généralisent la capture du geste manuel à celle du geste corporel. Ces dispositifs sont typiquement des *dispositifs à immersion* comme on les a déjà caractérisés plus haut, fonctionnant exclusivement en capteurs. Mais leurs constructeurs sont préoccupés depuis quelques temps, de les rendre bidirectionnels, c'est-à-dire de leur associer un "retour".

Le gant à retour tactile de Advanced Robotic Research Ltd (ARRL)

Un palliatif intéressant aux défauts du gant de données utilisé pour des manipulations d'objets virtuels est l'adjonction de dispositifs produisant la sensation du toucher. Le principe mis au point par Advanced Robotic Research Ltd (ARRL) dirigé par Robert Stone près de Manchester en Grande Bretagne, en collaboration avec la société Airmuscle Ltd - le *Teletact* - est étonnant. Il utilise 20 ballonnets situés au contact de toutes les phalanges et de la paume de la main. Chaque ballonnet est relié à un compresseur commandé par l'ordinateur et produit localement une pression engendrant la sensation d'un contact. Pour fournir les bonnes pressions aux bonnes phalanges et au bon moment en fonction de la manipulation, un dispositif symétrique d'"apprentissage" a été mis au point. Il s'agit d'un gant identique, mais où les ballonnets sont remplacés par autant de capteurs de force. La phase d'apprentissage s'effectue alors dans une opération de manipulation d'un objet réel, pendant laquelle on enregistre les informations de force de contact qui seront ensuite restituées à l'identique lors de la situation de manipulation de l'objet virtuel. Il n'y a pas à proprement parler de modélisation des objets et de leur interaction avec l'opérateur, il s'agit d'un relevé phénoménologique ad'hoc, pour chaque cas d'espèce.

Il est bien clair que l'information tactile renvoyée ici, n'est à nouveau qu'une *information*, c'est-à-dire que toujours rien n'empêche de traverser ou écraser les objets. En fait, comme je l'ai déjà souligné beaucoup plus haut, il ne s'agit pas d'un retour d'effort, mais d'un retour tactile. Il s'adresse à la perception tactile cutanée et non à la perception tactilo-proprio-kinesthésique.

L'exosquelette de la main de EXOS

Le "Dexterous Hand Master" de la société EXOS à Burlington dans le Massachusetts est un exosquelette articulé qui se monte sur la main. Les variations angulaires de chaque paire de phalanges articulées sont mesurées par un capteur angulaire monté à l'articulation des "exophalanges"⁶ correspondantes. La précision de la capture est très bonne mais le dispositif, bien que très léger (environ 30 g par degré de liberté), est une entrave importante au naturel des évolutions de la main. Une particularité intéressante de ce dispositif est la nature du retour tactile : il est réalisé par un actionneur vibrotactile, c'est-à-dire en fait une petite bobine de haut-parleur qui fonctionne à 250 Hz, fréquence à laquelle la peau présente un seuil de sensibilité minimum aux vibrations. Les modèles qui contrôlent le retour tactile sont, comme précédemment assez simplistes. Ce sont des modèles fonctionnels basés sur la tabulation de l'intensité du stimulus en fonction de la distance. L'échantillonnage des données du mouvement se fait à 60 Hz.

Notons qu'il ne s'agit toujours ici que de retour tactile et non tactilo-proprio-kinesthésique. Les dispositifs que je vais présenter maintenant, par contre, se caractérisent en premier lieu par la prise en charge du retour d'effort.

Le Manche de Commande à Retour d'Effort de MATRA-ESPACE

Développé récemment par MATRA-ESPACE, le "Robotop", manche de commande à retour d'effort est une poignée de pilotage à 3 degrés de liberté dotée de capteurs de force et de moteurs. Il permet de percevoir des "champs de force" qui "canalisent" le geste de l'opérateur et l'assistent dans des opérations telles que la génération de trajectoire ou, plus généralement lui permettent de percevoir des formes en mode tactilo-kinesthésique. Ce dispositif a été construit pour des applications telles que la téléopération d'un robot manipulateur où l'on retrouve le concept bras-maître/bras-esclave, ou l'interaction avec un environnement virtuel généré par ordinateur. Le conditionnement du retour d'effort se fait par programmation dans la station informatique à laquelle le manche est relié. Celle-ci reçoit les signaux des capteurs de force et fournit les signaux de retour à un étage de puissance qui commande les moteurs.

Il entre dans la catégorie des *dispositifs à vis-à-vis* et présente à ce titre des performances significatives en termes de retour d'effort et de perception tactilo-kinesthésiques. L'ambitus de déplacement spatial est d'une dizaine de cm sur lesquels peuvent être développées des forces de l'ordre de 25 N (2,5 Kgf). Les modèles pour le retour d'effort sont des modèles "fonctionnels", c'est-à-dire basés sur des profils types de fonctions de transfert Force-Position ou Force-Vitesse. On y trouve cependant quelques notions métaphoriques telles que "forces élastiques", "profils de raideur", "frottement visqueux", "champ de potentiel attractif ou répulsif", etc.

Le "Compact Master Manipulator" de l' Institute of Engineering Mechanics à Tsukuba

⁶ L'exosquelette, d'une manière générale, est une structure articulée comportant sensiblement les mêmes segments et les mêmes articulations que le squelette du membre concerné, mais avec une distortion géométrique qui permet de le monter en "enveloppe" du membre en question.

Ce dispositif, construit par Hiroo Iwata à l'Université de Tsukuba au Japon, entre lui aussi dans la catégorie *à retour d'effort* par l'intensité significative des forces de rétroaction qu'il produit, de l'ordre de 3 Kgf. Il est singulier par l'architecture mécanique mise en œuvre, comportant un système de pantographes sur lequel est montée une plateforme triangulaire permettant alors 6 degrés de liberté rétroactifs pour les déplacements de la main. Trois autres degrés de liberté, non rétroactifs s'ajoutent, à travers un système de préhension de la paume, de l'index puis des trois doigts inférieurs.

Associé à un poste de travail avec visualisation tridimensionnelle et un astucieux système de miroir à 45°, l'opérateur voit une image virtuelle de sa main, reconstruite par l'ordinateur, à l'emplacement de sa main réelle, qui reproduit ses propres mouvements. L'opérateur "plonge" ainsi sa main dans l'espace virtuel, à la manière des autres systèmes de RV⁷. L'échantillonnage du mouvement est peu rapide (de l'ordre de 10 Hz).

Avec la puissance significative qu'il propose en rétroaction sur 6 degrés de liberté, ce système est intéressant. La main est toutefois contrainte dans un espace relativement restreint, ce qui lui enlève les avantages des dispositifs *à vis-à-vis* sans lui donner les avantages des dispositifs *à immersion*.. Par ailleurs, la faible fréquence d'échantillonnage le range à nouveau du côté des systèmes de contrôle, plus que de celui des véritables systèmes de production de perceptions tactilo-proprio-kinesthésiques.

Le ©Clavier Rétroactif Modulaire de l'ACROE

L'équipe de l'ACROE, que je dirige avec Annie Luciani à l'Institut IMAG (Informatique et Mathématiques Appliquées) à Grenoble travaille depuis de nombreuses années sur la question du geste dans la communication homme/machine et en particulier sur les dispositifs à retour d'effort. C'est la raison pour laquelle, on l'aura compris, cet exposé témoigne d'une sensibilité et d'une exigence particulières quant aux performances nécessaires pour un retour d'effort s'adressant véritablement au sens tactilo-proprio-kinesthésique. Notre position, sur ce point, comme je l'ai déjà mentionné, est qu'il n'est pas possible de tout obtenir et que, dans l'état actuel de la technologie, il faut faire certains choix avec rigueur et détermination. En particulier, nous ne croyons pas que le véritable retour d'effort soit possible autrement qu'avec les systèmes à *vis-à-vis*. Aussi nous sommes-nous dès le départ orientés vers de tels systèmes et, en corrélation (mais également pour d'autres raisons) vers une approche spécifique de la communication homme/machine dans laquelle il est hors de question, sous-prétexte de le mettre en relation avec le monde virtuel de l'ordinateur, de couper l'opérateur du monde réel.

Le ©CLAVIER RETROACTIF MODULAIRE (CRM) de l'ACROE (CADOZ, LISOWSKI, FLORENS, 1988, 1989, 1990) est l'aboutissement de réflexions, de réalisations et d'expérimentations commencées vers 1974. Achievé en 1988, il fait suite à deux autres systèmes à retour d'effort, le premier construit en 1978 (FLORENS, 1978), le second en 1981 (CADOZ, FLORENS, LUCIANI, 1981). Partant de la morphologie de base d'un clavier de piano, dont il reprend la distance standard inter-touches (13,75 mm), il propose deux solutions au problème de la généralité des applications : la première en ce qu'il est modulaire, c'est-à-dire que l'on peut le constituer par extension avec autant de degrés de liberté que l'on souhaite, sans limitation de principe, la seconde en ce que sa morphologie est transformable. Un principe de montage mécanique simple permet en effet, selon l'usage, d'en faire un clavier type piano (avec la position et la mise en relief standard des touches blanches et des touches noires), ou un ensemble de sticks uni- bi-

⁷ RV : "Réalités Virtuelles" pour les intimes.

ou tri-dimensionnels, ou encore une pince de manipulation à N degrés de liberté. Mais la particularité principale de ce dispositif est que chacun des degrés de liberté est capteur et moteur. L'adoption du principe du *vis-à-vis* nous a permis très tôt de travailler dans une perspective de puissance motrice importante sous faible encombrement. La technologie spécifique du moteur mise au point ici, le *©Moteur en Tranches* permet des performances importantes : environ 8 Kgf de retour d'effort au point de manipulation, un temps de réponse du moteur à la consigne de force en entrée de l'ordre de 10 ms, pour un encombrement en largeur de 13,75 mm par degré de liberté (16 degrés de liberté dans la version actuelle), et un débattement de 60 mm en profondeur.

Ce dispositif est par ailleurs associé à un système de modélisation et de simulation d'objets physiques en temps-réel, le système CORDIS-ANIMA (CADOZ, 1979; CADOZ, LUCIANI, FLORENS, 1981, 1984, 1989; FLORENS, CADOZ, 1984, 1991) et la programmation du retour d'effort se fait non pas à partir de modèles phénoménologiques ou fonctionnels, mais de modèles physiques. Le point de vue est intégralement métaphorique : l'ordinateur simule des objets physiques aux propriétés dynamiques complexes, fines et très précisément définies. Par le moyen du CRM, l'opérateur ne fait pas que recevoir des informations tactiles, il touche, saisit, manipule des objets physiques virtuels. Le réalisme des sensations, qui fait que l'on peut véritablement parler non seulement de contrôle, mais de synthèse des sensations tactilo-proprio-kinesthésiques, tient aussi au fait que pour ces simulations, la fréquence d'échantillonnage (en mesure des positions comme en restitution du retour de force) est de l'ordre de 300 à 1000 Hz. Ces grandeurs sont par ailleurs impératives si l'on veut s'affranchir des problèmes d'instabilité de manière satisfaisante.

Bien évidemment, la contre-partie est que ce dispositif fonctionne en *vis-à-vis*, c'est-à-dire qu'à aucun moment l'opérateur ne peut envisager d'"entrer corps et âme" dans le monde virtuel de l'ordinateur. L'ordinateur et ses périphériques lui apparaissent comme un objet parmi les objets du monde environnant, avec lequel l'opérateur peut entrer en relation, mais nous avons tendance à penser que ceci lui est plutôt salutaire.

3.3. Paradigmes Artificiels

Les dernières remarques me conduisent justement à proposer un état comparatif des différents systèmes non plus seulement en termes de performance ou d'adéquation à la tâche, mais plutôt en fonction de leur "signification". Chaque attitude manifeste en effet, la plupart du temps implicitement, une façon de concevoir la relation entre les divers protagonistes - l'opérateur, un interlocuteur (ou environnement humain), la tâche (ou environnement physique) - les uns et les autres pouvant, selon le cas se présenter comme réels ou virtuels (ou artificiels).

Dans l'approche de Myron Krueger par exemple, l'homme lui-même est virtualisé, et son double est envoyé, avec éventuellement quelques collègues, dans un monde également virtuel pour interagir avec eux, dans ce monde là. De cette interaction il émane quelques manifestations que l'observateur réel considère et qui lui permettent d'influencer les évolutions des protagonistes virtuels.

Dans les "réalités virtuelles" avec gant, exosquelette, HMS et autres dispositifs à immersion, l'opérateur reste réel, il n'a pas de double comme dans le cas précédent, mais il est plongé intégralement et sans recours dans l'espace virtuel. Il peut y rencontrer des objets virtuels, mais aussi des personnages qui, de leur propre point de vue sont restés réels, mais lui apparaissent alors comme virtuels.

Dans le cas des systèmes à immersion partielles comme le gant sans HMS, en face d'un écran, ou le *Compact Master Manipulator* de Hiroo Iwata, l'opérateur est moitié homme / moitié robot, enfin plutôt moitié réel / moitié virtuel. Seules certaines parties de son corps sont virtualisées et plongées dans l'espace virtuel, mais il reste solidaire de ces parties virtuelles.

Enfin, dans le premier (le projet GROPE) comme dans le dernier (le CRM de l'ACROE), l'opérateur reste réel et le monde aussi. Dans ce monde réel, il y a des objets qui sont virtualisés, éventuellement réellement accessibles à la vision ou au toucher etc. par d'autres opérateurs réels, et l'opérateur réel entre en contact manipulateur, voir sensori-moteur et multisensoriel avec ces objets.

BIBLIOGRAPHIE

- ATKINSON, W.D., BOND, K.E., TRIBBLE, G.L., WILSON, K.R. - Computing with feeling - Comput. and Graphics, Vol 2 - 1977
- BATTER, J.J. and BROOKS, F.P., Jr. - GROPE-I - A computer display to the sense of feel - *Information Processing, Proc. IFIP Congress 71*, 759-763. 1971
- BEAUBATON, D. - Le contrôle de la main : planification du geste et assistance sensori-motrice - Institut de Neurophysiologie et de Psychophysiologie de Marseille 1980
- BEAUBATON, D. - Contrôle proactif et rétroactif de la motricité - Thèse à la faculté des sciences de Luminy (Aix-Marseille II) 1983
- BEEK, P.J. - Juggling Dynamics - Free University Press, Amsterdam 1989
- BEJCZY, A.K. and SALISBURY, J.K. - Controlling Remote Manipulators Through Kinesthetic Coupling - Computer in Mechanical Engineering, July 1983, pp. 48-60. 1983
- BERTHOZ, A. - Rôle de la proprioception dans le contrôle de la posture et du geste - Contribution de quelques récepteurs à la détection des mouvements - 1975
- BOFF, K. KAUFMAN, L. THOMAS, J. - Handbook of Human Perception and Human Performance - volume I and II, John Wiley and sons, N.Y., 1986
- BRETON, P. - Histoire de l'Informatique - Edition la Découverte, Paris, 1987
- BROOKS, F.P., Jr. - Grasping reality through illusion : Interactive graphics serving science" - 5th Conf. on Comput. and Human Interact. - CHI'88. 1988
- BROOK, F.P., OUYOUNG, M., BATTER, J.J., KILPATRICK, J. - Project GROPE ; Haptic Displays for Scientific Visualization - Computer Graphics, Vol. 24, N° 4, August 1990
- CADOZ, C. - Synthèse sonore par simulation de mécanismes physiques vibratoires - Thèse INPG Grenoble France - 1979
- CADOZ, C., LUCIANI, A., FLORENS, J.L. - Synthèse musicale par simulation des mécanismes instrumentaux transducteurs gestuels rétroactifs pour l'étude du jeu instrumental - Revue d'acoustique N°59 - 1981
- CADOZ, C., LUCIANI, A., FLORENS, J.L. - Responsive input devices and sound synthesis by simulation of instrumental mechanisms : the CORDIS system - Computer Music Journal - N°3 - 1984 reprint in
- CADOZ, C., LUCIANI, A., FLORENS, J.L. - Responsive input devices and sound synthesis by simulation of instrumental mechanisms : the CORDIS system - in Music Machine Curtis Road ED. MIT Press, Cambridge Mass 1989
- CADOZ, C., LISOWSKI, L., FLORENS, J.L. - Clavier rétroactif modulaire - Computer Music Journal - Vol 14, N°2 - 1990
- CADOZ, C. - Commande instrumentale et synthèse sonore ; simulation de l'univers instrumental - in "La Machine perceptive - les moyens de communication entre l'ordinateur et l'homme" - 12 ièmes Journées francophones sur l'informatique - Grenoble - Jan. 1990
- COOKE, J.D. - Dependence of human arm movements on limb mechanical properties - Barin Research, 165 p.366-369 - 1979
- DANNENBERG, R.B. - A Gesture Based User Interface Prototyping System - ACM 1989
- DORGUEUILLE, DR. - Essai d'analyse anatomo-physiologique de la technique de clavier - Bulletin du Groupe d'Acoustique Musicale, juin 1967
- EKMAN, P. et FRIESEN, W.V. - Hand movements - The journal of communication, 22 December, 353-374 - 1972
- FANTAPIE, H.C. - L'analyse de la partition dans la pratique du chef d'orchestre - Revue d'Analyse Musicale n°10, p.26-30 - 1er trimestre 1988
- FEYEREISEN, P. - Gestures and speech, interaction and separations : A reply to McNeil (1985) - Psychological Review, 94, 493-498 - 1987
- GIBSON, J.J. - The senses considered as perceptual systems - Boston, Houghton Mifflin Company, 1966
- FLORENS, J.L. - Coupleur gestuel rétroactif pour la commande et le contrôle de sons synthétisés - Thèse INPG, Grenoble, France - 1978
- FLORENS, J.L., CADOZ, C. - Geste instrumental, ordinateur, transducteurs gestuels rétroactifs - 10ème ICMC, p.59-61, IRCAM, 1984
- FLORENS, J.L., CADOZ, C. - The physical Model, Modelisation and Simulation Systems of the Instrumental Universe - In representation of Musical Signals - G. De Poli, A. Picciali, C. Roads, Ed. MIT Press, 1991